REMARKS

Claims 20-23 and 25-29 are now pending in the application. Claims 1-19 and 24 have been cancelled. Claims 20-23 and 25-26 are currently amended. Claims 28 and 29 have been added. The Examiner is respectfully requested to reconsider and withdraw the rejection(s) in view of the amendments and remarks contained herein.

REJECTION UNDER 35 U.S.C. § 112

Claims 19 and 26 stand rejected under 35 U.S.C. § 112, second paragraph, as failing to comply with the written description requirement. This rejection is respectfully traversed.

Claim 19 has been cancelled and Claim 26 has been amended to comply with the written description requirement.

REJECTION UNDER 35 U.S.C. § 102

Claims 19-27 stand rejected under 35 U.S.C. § 102(b) as being anticipated by Herz (U.S. Pat. No. 6,029,195). This rejection is respectfully traversed.

1. Claim 20

The feature of the system recited in the amended claim 20 resides in including: clustering means for classifying the plurality of documents into a plurality of clusters; cluster term label preparation means for selecting one or more terms as a label of the cluster for each of the plurality of clusters; document retrieval means for retrieving a document satisfying a retrieval condition input by the user among the plurality of

documents; and interface means for presenting the retrieved document together with the label of the cluster and the rest of documents belonging to the cluster as retrieval results. Fig. 3 of the present application illustrates an example of the presented retrieval results. In Fig. 3, the term label of the cluster is presented as "Confectionery, snack, cheese" in the upper portion at the "cluster label" column. The block of similar documents is labeled as a term label. Therefore, the user can easily grasp the retrieval results (see page 12, lines 23-25 of the specification).

The procedure of term label preparation of the present invention is described from page 14, line 4 to page 15, line 4, and will be described in detail with reference to Appendix Fig. 1. Appendix Fig. 1 illustrates an example of a procedure of term label preparation of the present invention. First, the documents 1 to 3 are classified into a cluster by the clustering means. Next, the term scores of each term included in the documents belonging to the cluster are counted. Note that only nouns are selected as the terms. Then, the terms, e.g., three terms, having high term scores are selected as a label. The terms having high term scores are ones which frequently appear in the documents belonging to the cluster. The inventors, through the experiments, found that when the terms having high term scores are presented as a label, the user can easily grasp the retrieval results. The present invention achieved the significant effect as a label of the cluster to facilitate the user's grasp of the retrieval results. Therefore, the present invention cannot be easily anticipated by one skilled in the art.

On the other hand, Herz discloses selecting the term having large TF/IDF score as a label (see col. 71, line 47). In applying this method to an example shown in Appendix Fig. 1, when the term "alligator" which hardly appears in all the documents

(i.e., the IDF score is high), the TF/IDF score increases because of the constant TF score. Thus, even when the term "alligator" appears only in one document of the cluster, the term is selected as a label. Hence, Herz exhibits a different result from the present invention. It is inappropriate to select as a label the term which does not commonly appear in the cluster.

Moreover, the invention recited in the amended claim 20 limits the clustering procedure. The procedure of calculating a distance between the feature vectors of the documents is described in page 11, lines 5-24 of the specification. In the present invention, a similarity ratio representing the distances between feature vectors is defined. The effect of the similarity ratio will be described in detail with reference to Appendix Fig. 2.

To define a distance between the feature vectors, Euclidean distance and cosine are usually used. In Appendix Fig. 2, the range of distance equal to or less than the distance between the feature vector OA and the feature vector OB is presented as a circle having the radius of AB and the center point of A in the Euclidean distance, as an area surrounded by Ray OC and Ray OD in cosine, and as Quadrangle EFGH in similarity ratio.

In the Euclidean distance, the longer the vectors are, the narrower the area of the relatively similar vectors is. Therefore, as the document is long, the similarity becomes low even in the documents having the similar terms.

In the cosine, the similarity is determined only by the direction, regardless of the length of the vectors. Thus, despite of the length of the document, the similarity of the terms which frequently appear becomes high.

In the similarity ratio which is employed in the present invention, the area of the similarity is narrower in the range of short vectors and on the other hand, it is wider in the range of long vectors. Accordingly, the similarity is high only in the case where both the length of the document and the appearing terms are similar. The distance is appropriate for the case of clustering the documents having various lengths.

None of the references discloses using the distance of the present invention in clustering the documents, and hence the present invention is not identical to the invention disclosed in the references.

2. Claim 21

The feature of the system recited in the amended claim 21 resides in including: clustering means for classifying the plurality of documents to be retrieved into a plurality of clusters; cluster sentence label preparation means for selecting one sentence as a label of the cluster; document retrieval means for retrieving a document satisfying a retrieval condition input by the user among the plurality of documents; and interface means for presenting the retrieved document together with the label of the cluster and the rest of documents belonging to the cluster as retrieval results. Fig. 3 of the present application illustrates an example of the presented retrieval results. In Fig. 3, the sentence label of the cluster is presented as "Watery food (jelly, pudding, yogurt)..." in the lower portion at the "cluster label" column. The block of similar documents is labeled as a sentence label. Therefore, the user can easily grasp the retrieval results (see page 12, lines 23-25 of the specification).

The procedure of sentence label preparation of the present invention is described from page 15, line 14 to page 16, line 15, and will be described in detail with reference to Appendix Fig. 3. Appendix Fig. 3 illustrates an example of a procedure of sentence label preparation of the present invention. First, the clustering means classify the documents 1 to 3 into a cluster. Next, the term scores of each term included in the documents belonging to the cluster are counted. Note that only nouns are selected as the terms. Then, the sum of the term scores of the terms included in each of the documents belonging to the cluster is obtained. The sentence having the largest sum is selected as a label. Thus, the sentence having high frequency of the term which appears commonly in the documents belonging to the cluster is selected as a label, and the sentence is a representative sentence in the cluster. The inventors, through the experiments, found that the aforementioned procedure facilitates the user's grasp of the retrieval results.

Herz teaches using the term row as a label of the cluster, but fails to disclose using the sentence as a label.

Note that the amended claim 21 limits the clustering procedure as in the amended claim 20, and thus the same argument on the clustering procedure is applicable to the amended claim 21.

3. Claim 22

The procedure of term label preparation by the cluster label preparation section is described in page 14, line 4 to page 15, line 4 of the specification. As mentioned above, the inventors found that the sentence in which the sum of the term scores is

largest is selected as a label of the cluster, thereby facilitating the user's grasp of the retrieval results.

In the case where a plurality of sentences having the same sum of the term scores exist but only one sentence label is desired, it is necessary to further select the terms in the plurality of sentences having the same sum. The inventors found that in the aforementioned case, the sentence having the smallest number of characters is selected among the plurality of sentences having the same sum, thereby facilitating the user's grasp of the retrieval results.

According to the procedure recited in the amended claim 22, the significant effect of the cluster label preparation can be achieved in addition to the effect obtained from the invention recited in the amended claim 21. Hence, the present invention cannot be easily anticipated by one skilled in the art.

4. Claims 23 and 25

The features of the system recited in the amended claims 23 and 25 reside in including: the clustering means for classifying the documents to be retrieved into a plurality of clusters; the cluster label preparation means for generating a cluster label containing the sentences for each of the plurality of clusters; the document label preparation means for preparing a document label representing the contents of the document; the document retrieval means for retrieving a document satisfying a retrieval condition input by the user; and interface means for presenting as retrieval results the retrieved document together with the cluster label of the cluster to which the retrieved document belongs, the rest of documents belonging to the cluster, and the document

labels which are associated with each of the retrieved document and the rest of documents. Fig. 3 of the present application illustrates an example of the presented retrieval results. In Fig. 3, the cluster label is presented at "cluster label" column, and the document label is indicated by the underline in each document at "documents" column. The cluster includes blocks of similar documents and is labeled, and the feature portions of each document belonging to the cluster are underlined. Therefore, the user can easily grasp the retrieval results (see page 12, lines 23-25 of the specification).

As mentioned above, the cluster label enables the user to easily grasp the contents of the cluster including the documents. The inventors, moreover, found that in the case where the user focuses one cluster, the user may need the document label representing the contents of the documents belonging to the cluster. The procedure of document label preparation is described from page 17, line 4 to page 18, line 7 of the specification. In the cluster label, the term or the sentence representing the documents in the cluster is selected, while in the document label, the distinction among the similar documents in the same cluster must be highlighted. The inventors, through various experiments, found that the selection of the document having the largest sum of TF/IDF values facilitates the user's grasp of the retrieval results.

Herz teaches the cluster label, but fails to disclose any document label in the cluster.

Note that the amended claims 23 and 25 limit the clustering procedure as in the amended claim 20, and thus the same argument on the clustering procedure is applicable to the amended claims 23 and 25.

5. Claim 26

The feature of the system recited in the amended claim 26 resides in including: clustering means for classifying the plurality of answer documents into a plurality of clusters; question document retrieval means for retrieving a question document conforming with a user question input by the user among the plurality of question documents; and interface means for presenting the retrieved question document and the answer document associated with the question document together with the rest of answer documents included in the cluster to which the answer document belongs, as retrieval results. The construction of the system is described in Embodiment 2 of the specification, and will be described in detail with reference to Appendix Fig. 4.

In Appendix Fig. 4, assume that the user retrieves the question document ③ although the user wants to obtain the answer document ①. First, the answer document ② associating to the question document ③ is obtained. The answer document ① belonging to the same cluster as the answer document ② is also presented as the retrieval result. Thus, the user can obtain the desired answer document.

In other words, since the similar answer documents are classified into the same cluster, the user can obtain the desired answer document as a retrieval result even when the desired answer document is not one associated with the question document retrieved from the user's question. This is because when the desired answer document is similar to the answer document associated with the retrieved question document, the block of the answer documents is presented as a cluster.

Therefore, the system of the present invention can present the most appropriate answer document with respect to the user's various questions in a fewer association of the question documents with the answer documents.

In fact, in the case of answering the user's questions, fewer answer variations are needed as compared with the user's question variations. Thus, the relationship between the user's question and the answer is multitude to 1. The conventional question/answer system associates the question documents with the answer documents one to one. Therefore, when a large number of pairs of question and answer documents are prepared to deal with various questions, the similar answers come up for the answer documents and the desired answer document cannot be presented. Particularly, the question/answer system storing examples of the actual questions and answers often registers pairs of questions and answers having the overlapping contents, and thus the aforementioned problem significantly comes up. To avoid the problem, if pairs of questions and answers having the overlapping contents are reduced, few questions can lead to appropriate answers. In addition, the administrator's burden increases. Alternatively, if the examples of questions and answers are arranged so as to meet the relationship of multitude to 1, the administrator's burden is all the more heavy.

In the present invention, the answers are not associated with the questions one to one, but one answer is associated with a plurality of questions (see page 21, lines 4-17 of the specification). In addition, the question documents are retrieved to determine whether the similar documents exist with respect to the questions input by the user. The answer documents are brought together into a cluster by the clustering means.

Thus, the relationship between the questions and the cluster of answers is multitude to one. According to the present invention, many variations of questions are stored in the system to deal with inquiries, but the answer list is not occupied with the similar answers. As a result, the user can easily grasp the retrieval results (see page 12, lines 23-25 of the specification). Moreover, the relationship of the questions and the cluster of answers, i.e., multitude to one, is automatically formed, and thus the administrator's burden does not increase.

Herz teaches an application in which the document having a pair of question and answer is retrieved, but merely discloses retrieving the document having the question and the answer one to one.

Note that the amended claim 26 limits the clustering procedure as in the amended claim 20, and thus the same argument on the clustering procedure is applicable to the amended claim 26.

6. Claims 27 and 28

The features of the invention recited in claims 27 and 28 reside in that the user selects an answer document among the answer documents in the cluster presented as retrieval results, and the document of the user question is newly stored in the document storage means in association with the selected answer document. According to the configuration, when the similar user question is input in the next and subsequent times, the answer document which is newly stored as the retrieval result can be retrieved. Since an appropriate association between the question and the answer is automatically

updated, the information retrieval system can provide the more appropriate answer without putting a burden upon the administrator.

In the invention disclosed in Herz, the hierarchies of menu of clustering are updated in accordance with feedback from users. However, Herz fails to disclose updating the document(s) stored in the document storage means.

7. Claim 29

To deal with the same case as in the above Claim 22, the inventors found that a sentence the head of which is located nearest to the beginning of the document is selected among the plurality of sentences having the same sum, thereby facilitating the user's grasp of the retrieval results.

According to the procedure recited in the new claim 29, the significant effect of the cluster label preparation can be achieved in addition to the effect obtained from the invention recited in the amended claim 21. Hence, the present invention cannot be easily anticipated by one skilled in the art.
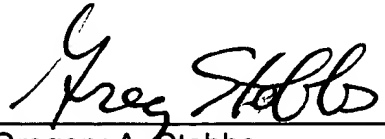
CONCLUSION

It is believed that all of the stated grounds of rejection have been properly traversed, accommodated, or rendered moot. Applicant therefore respectfully requests that the Examiner reconsider and withdraw all presently outstanding rejections. It is believed that a full and complete response has been made to the outstanding Office Action, and as such, the present application is in condition for allowance. Thus, prompt and favorable consideration of this amendment is respectfully requested. If the

Examiner believes that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at (248) 641-1600.

Respectfully submitted,

Dated: Apr 20, 2005

By: _Greg Stobbs_____

Gregory A. Stobbs
Reg. No. 28,764

HARNESS, DICKEY & PIERCE, P.L.C.
P.O. Box 828
Bloomfield Hills, Michigan 48303
(248) 641-1600

GAS/kk